IADIS International Journal on Computer Science and Information Systems
Vol. 4, No. 1, pp. 64-74
ISSN: 1646-3692

A GEO-SPATIAL DECISION-SUPPORT SYSTEM FOR REGIONAL AIR QUALITY MANAGEMENT

El MOHAJIR Mohammed, OUBENAALLA Youness
Department of Mathematics and computer sciences, Faculty of sciences Dhar Mahraz, BP 17.96, Fez, Morocco

ABSTRACT

This paper describes the geo-spatial multidimensional database architecture we built to construct a regional air quality decision-support system. The system retrieves data from three main categories of system sources: a numerical atmospheric model, in-situ detectors and satellites. We worked on an idealized case study for which we established spatial multidimensional OLAP cubes on specific measures and dimensions. This schema aims to provide original analysis like deviations of the concentration simulated by the atmospheric model per region, per date and per pollutant. Cartographic representation of the results is then achieved by using the SOLAP approach. Thematic maps are generated as a response to a particular query that the end user may submit. A range of options are offered to explore these charts in both space and time. Cartographic contours are not based on pre-defined images but are dynamically elaborated from the numerical coordinates stored in the spatial dimension. Qualities of the design and implementation are finally assessed according to recognized criteria established by the OLAP/SOLAP community as scalability, time response and data exploration intuitiveness.

KEYWORDS

Geo-Spatial Decision-support system, SOLAP, Regional air pollution, Numerical atmospheric model, In-situ captor, Satellite observations.

1. INTRODUCTION AND RELATED WORKS

Multidimensional structures are powerful in supporting efficiently the decision-making process as they provide pertinent and in-time responses to specific users' queries. They provide aggregate information witch is re-arranged in the most optimal database schema. OLAP (One Line Analytical Processing) structures are easier to establish and to query than transactional structures called OLTP (On Line Transactional Processing). Deficiencies in these approaches are filled in by SOLAP (Spatial OLAP) technology that offers cartographic options that are provided neither by the traditional GIS nor by OLAP tools (Bédard et al. 2006).

At the present time, the use of SOLAP technologies covers a large range of applications, for example, the projects conducted at ***"Centre de Recherche en Géomatique de l'Université Laval"***, related to forestry management, lorry driver networks, and environmental health (Bédard.et al. 2006). Other applications using SOLAP are developed at LIRIS to monitor the death rates for the different departments of France (Bimonte et al. 2006).

Within the framework of the monitoring of air pollution, some research works were performed like the work realized by (Taher et al. 2005) and (Richards et al. 2006).

In this paper, we use the SOLAP approach to develop a decision-support system for regional air pollution monitoring. The data collection in our case is heterogeneous as it is obtained from three main sources: data produced by a numerical atmospheric model, concentrations captured by in-situ detectors and observed by satellite. Within this framework, we will present the SOLAP architecture we defined as well as the data integration mechanisms used for extraction, analysis and exploration purposes. We will define specific queries that can be either spatial or spatio-temporal. The cartographic results will be illustrated too. We will also discuss the benefits and advantages of this representation for supporting the decision making process, analysts or any end user queries.

The outline of our paper will be as follows: We will first give details of our case study, the application domain and the data specifications that will need to be treated. We will then present the SOLAP architecture we elaborated within this framework. The details of the Geographical Information System implementation are provided in the section 4. Finally, we will discuss the obtained results by comparing them to existing works. Relevance and quality of the thematic charts are also determined on the basis of recognized standards and criteria that are relevant to any DSS (Decision Support System).

2. GENERAL FRAMEWORK AND CASE STUDY

Numerical atmospheric models run over a range of spatial scales from the global scale (several thousand kilometers) to the regional scale (100-200 Km) with resolutions from 1-2 Km to 100 Km. Numerical models are used to carry out many kinds of simulations which make them powerful research tools for studying atmosphere dynamics and processes. They can run with several horizontal and vertical resolutions, and can also involve complex representations of physical and chemical mechanisms of the atmosphere (Chimere 2005).

Numerical models start from prescribed sets of initial conditions for the domain of study. We aim to use a prediction numerical model in order to anticipate fatal air pollution picks that obviously have drastic impacts on human and environment.

A numerical model includes the following components:

- Meteorological numerical model of short time range prediction.
- Dispersion – diffusion numerical model representing horizontal and vertical advection of the chemical species.
- Chemical model incorporating reactions between the atmosphere constituents.

3. SYSTEM ARCHITECTURE

The data warehouse that forms the core of our geo-spatial DSS incorporates information that comes from three main sources: the numerical model, captors and satellites. The treatment of full non-formatted produced data is not considered in this work; we will only consider part of the produced data and will assume that it has the appropriate format. Therefore, the production data base will contain: pollutant's concentrations, the wind's components and temperature, over the whole three-dimensional application domain.

We chose for the development of our application a combination between the geographical information systems (GIS) and OLAP tools. This coupling process called SOLAP (Bédard et al. 2006) (Proulx et al. 2006) (Rivest et al. 2001) will allow a thorough spatial analysis by providing thematic maps as results to spatio-temporal queries, additional operations like aggregation and drilling of data can also be performed. The following schema (figure.1) presents the main components of the developed system with the corresponding data flow:

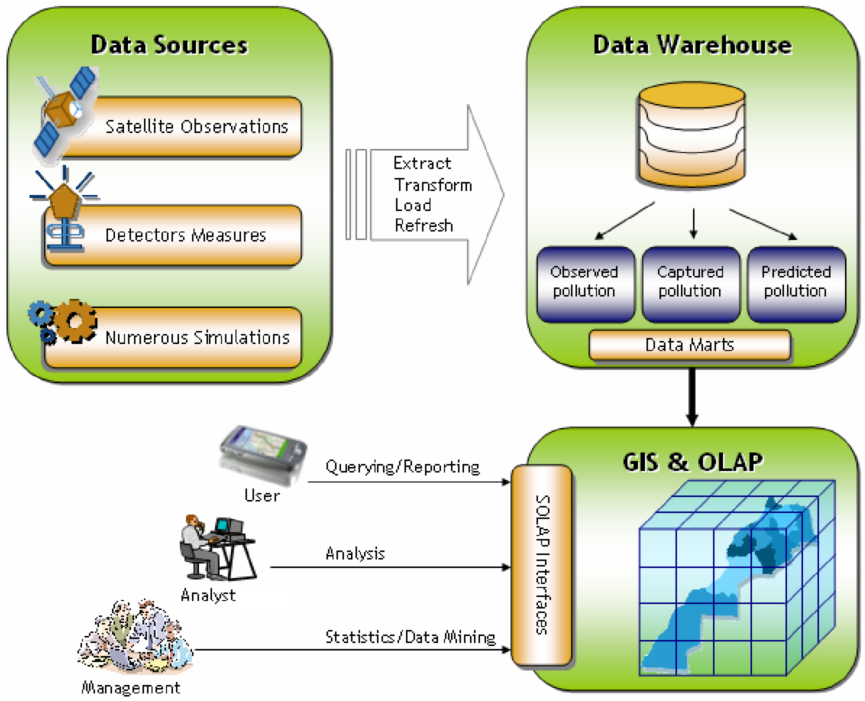

Figure 1. The overall architecture of the air pollution system.

The first component includes the system data sources; it shows heterogeneous information obtained from satellite observations, sensors measures and numerous prediction simulations. The second component is the environmental data warehouse. The data are extracted from transactional databases or flat files, transformed to multidimensional architectures like star, unified flake or constellation, and then loaded within data marts, created during the transformation process. The third component relates to the SOLAP architecture. The data is geo-referenced under spatio-temporal cubes. The last component of the system includes the user's interface which is used for queries, reports generation and data analysis. The interface offers different options based on the profile of the connected user like analysts, administrators or decision makers.

4. ENVIRONMENTAL DATABASE ARCHITECTURE

4.1 Description of the OLAP context

OLAP technologies allow the implementation of multidimensional databases. This approach is based on concepts like "*dimensions*" which represent the themes of interest and which often are organized hierarchically into different levels of granularity (ex: Pollutant, Time, Detector…), and "*measures*" which are numerical facts analyzed according to specific axis. For example, we can visualize the values of the NO2 pollutant per area and per date.

Each dimension is described by a set of "members" (ex: Semester1 of year 2006, Fez_Medina, combustible Pollutant…), the members of one level (different areas) may be aggregated to constitute the members of the next higher level (the City for example). The "facts" represent the subject of interest that relates specific measure to specific dimensions on a particular aggregation level (Chaudhuri et al.1997) (Kimball. 2002) (Vassiliadis. 1998) (Agrawal R et al. 1995).

The following OLAP schema depicts the overall multidimensional structure we will use in the context of our study:

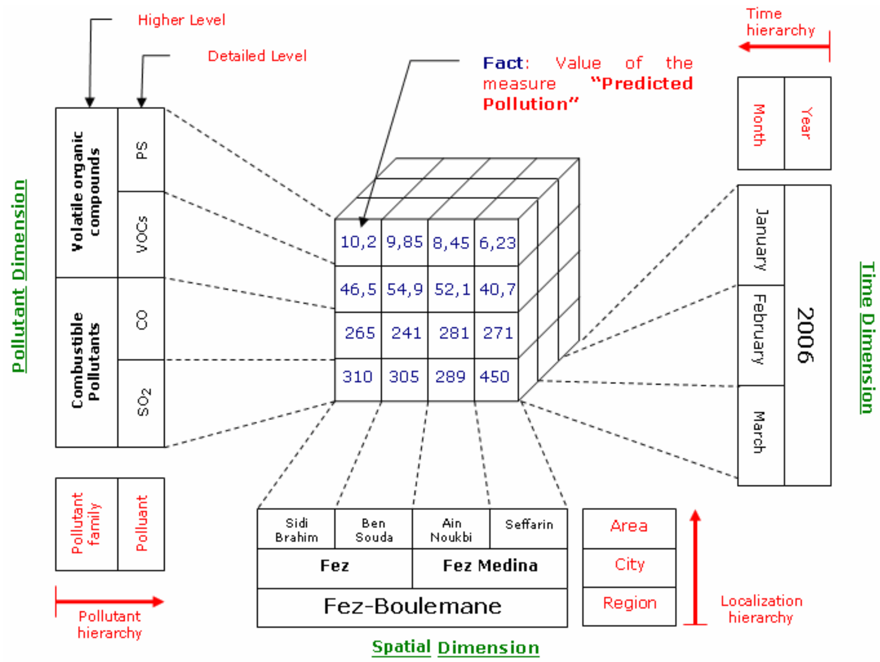

Figure 2. Example of an air pollution multidimensional cube presented in our system

4.2 Spatio-temporal granularities and OLAP operators

The OLAP technology offers efficient operations for navigating among "hierarchies" and "levels" and also interactivity in term of data exploration. Popular operations that are applied to multidimensional cubes are *Drill Down*, *Roll Up*, *Drill Across*, *Slice* and *Pivot*. *Drill Down* operation allows visualizing a more detailed level within a dimension (drilling down the Fez-Boulmane *Region* to obtain values in the detailed level C*ity*). Roll Up corresponds to a *group-by* on one of the dimensions (ex. rolling up all semesters to have the value of O_3 in a year). *Drill Across* consists to visualize another member or another dimension at the same level of detail. The *Pivot* operation consists in exchanging dimensions in order to modify the content of the analysis axis. The S*lice operation* consists in reducing the dimensionality of the data.

The following figure (figure.3) shows dimensions hierarchies we are using for the pollution multidimensional model.

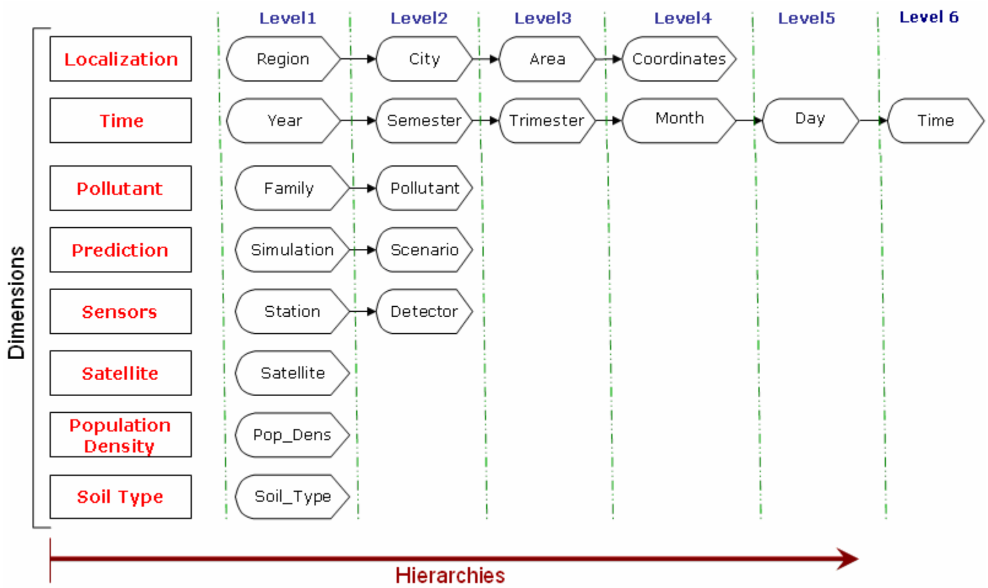

Figure 3. Hierarchies and granularities of the dimensions of the multidimensional schema.

In the present study, the themes of analysis considered are: the concentrations of pollutants captured by pollution sensors, those observed by satellites, and those simulated by the numerical atmospheric model.

5. GRAPHICAL INFORMATION SYSTEM AND SOLAP SPECIFICATIONS

5.1 GIS specifications

The Geographical Information Systems (GIS) are powerful tools to handle, query and navigate within geo-spatial databases (Bédard et al. 2006) (Franklin. 1992). A spatial data is often made of three parts: geometrical (ex. Line, polygon…), descriptive (ex. City_Name…) and calculated (ex. Sidi_Brahim_Global_Surface or Perimeter). The geographic information is represented either in *vector format* or *Raster format*.

5.2 SOLAP specifications

The SOLAP solution corresponds to the coupling of an OLAP server and a GIS tool. (Bédard et al. 2006) defines the SOLAP as "*a category of software that allows rapid and easy navigation within spatial databases and that offers many levels of information granularity, many themes, many epochs and many display modes synchronized or not: maps, tables and diagrams*".

SOLAP technology allows the dynamic creation of charts which are based on multidimensional data stored in the spatial cubes and offers the possibility of spatio-temporal exploration and analysis.

5.3 Data loading and interfaces

Data are extracted from hyper-cubes and geo-referenced on layer maps. The user uses different SOLAP operations like Rolling up or drilling down into *Areas* for the *localization* dimension for example. Thematic charts are obtained as a response to user's queries. They can also determine different aggregated values like Average, Max, Min…etc. The following figure (figure.4) shows an instance of the extraction and geo-referencing processes.

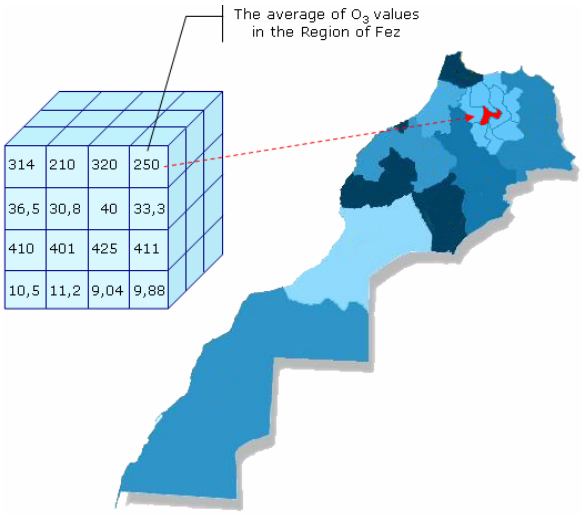

Figure 4 . Extracting multidimensional data and geo-referencing.

6. DATA EXPLORATION AND ANALYSIS

In terms of functionalities, our system offers a wide range of possibilities in exploring measures according to various axes of analysis. We give bellow some examples of how the cubes were deployed (ProClarity, 2006)..

6.1 Tabular format

Table.1 shows the average of pollutant values obtained by applying the aggregate operations; *Sum* and *Count* in a *day* per captor, then we *roll-up* the results to the *month* level and apply a second roll-up operator to the *semester* level.

Time	Detector	Pollutant	AVG
Semester1 / 2006	DET_SidiBrahim	O3	210.666
		CO	14946,333
		NO2	301.5
	DET_BenSouda	O3	215
		CO	15466,666
		NO2	297.66
	DET_AinNoukbi	O3	200.66
		CO	15622,25
		NO2	305.333

Table 1. The average of captured values of different pollutants per sensor in the first semester of 2006.

Table.2 shows the values of the pollutants for the period between 20/09/2006 to 22/09/2006, with roll up to the City level.

City	Time	pollutant	Station	P_Value
FEZ	20/09/2006	NO2	SIDIBRAHIM	400
			BENSOUDA	360
			DOUKARAT	157
	21/09/2006	O3	SIDIBRAHIM	180
			BENSOUDA	300
			DOUKARAT	150
	22/09/2006	SO2	SIDIBRAHIM	400
			BENSOUDA	268
			DOUKARAT	320

Table 2. Predicted values of different pollutants per Station from 20 to 22 September, 2006 in Fez City.

6.2 Diagrams

The diagram below shows the variation of different pollutant concentrations measured during 2006 in Fez, Rabat, Casablanca and Meknes.

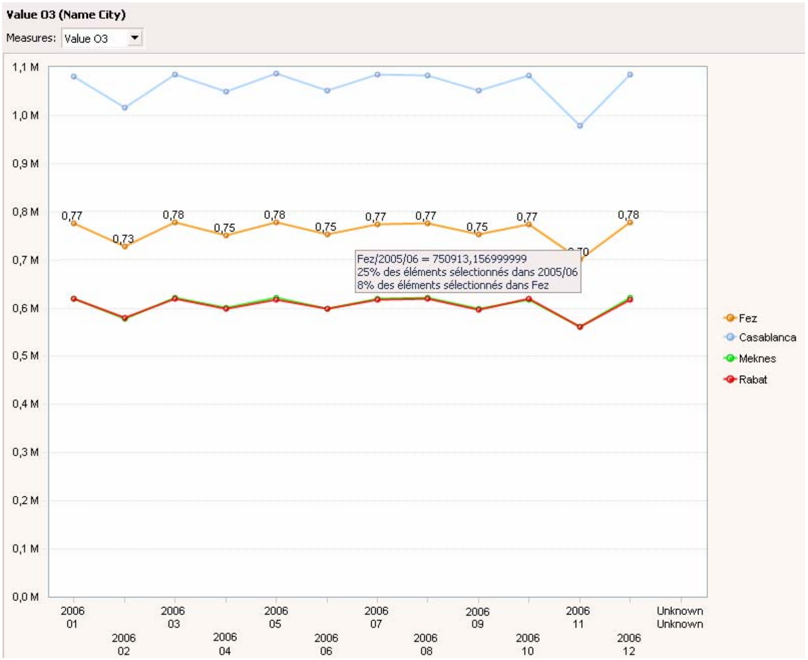

Figure 5. Pollutants concentrations measured during 2006 in various cities.

6.3 Thematic Maps

The SOLAP approach we adopted to query the geo-spatial data warehouse allows a cartographic representation of specific indicators on a selected area of the application domain. Different levels of thematic spatio-temporal maps can be generated as a result to user's application of the classical operators like drilling down or rolling up spatial dimension. The user can also analyse diverse aggregated values such as AVG, Min… The figure below (figure.6) shows an example of the extraction and geo-coding process to calculate the average of some pollutants concentrations in FEZ-BOULMANE Region. Spatial OLAP operations are also illustrated.

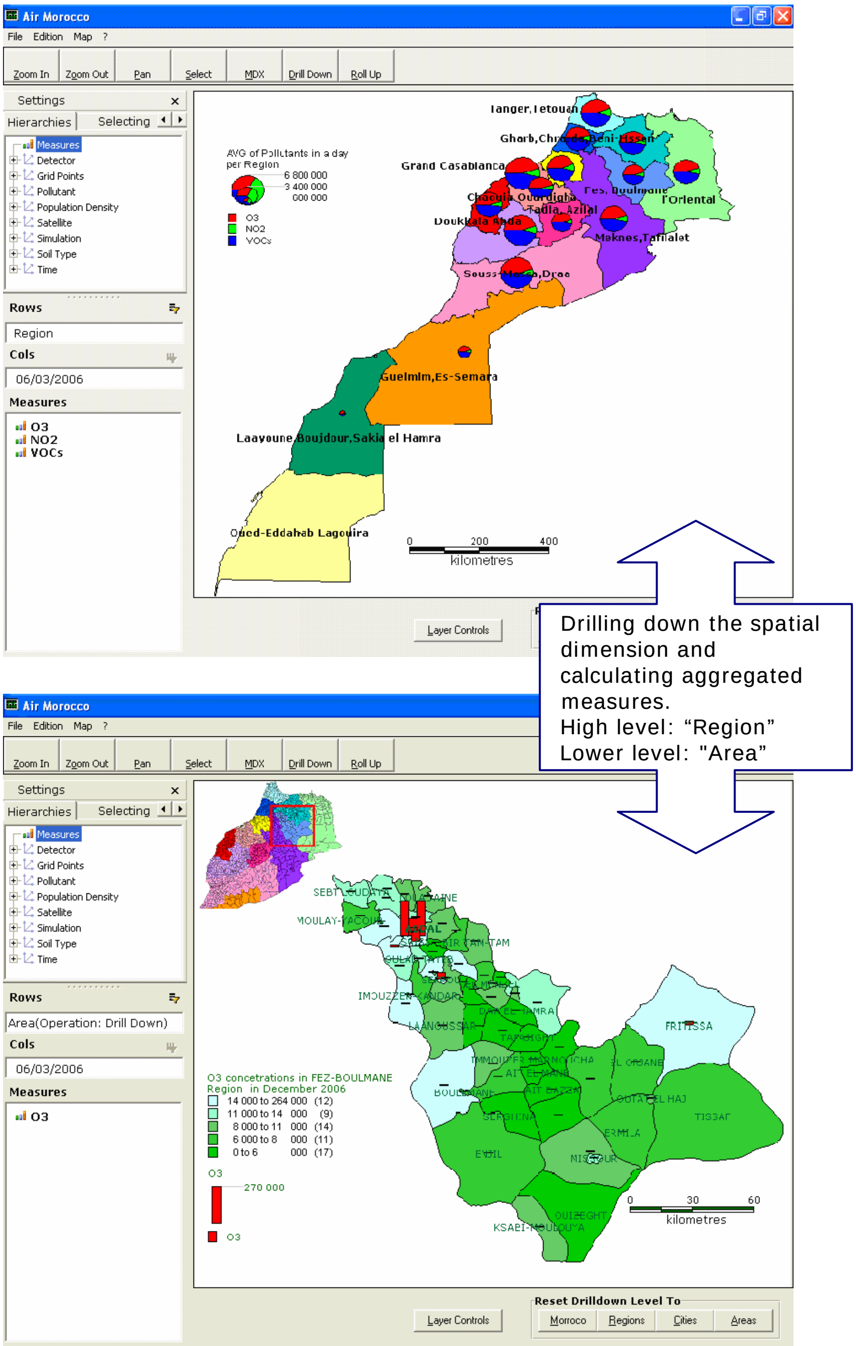

Figure 6. Drilling down the spatial dimension and thematic maps generation.

7. CONCLUSION AND FUTURE WORK

This work was completed within the framework of a development of a geo-spatial decision support system for regional air quality management. The application domain covers all Moroccan regions and cities. The integration of the warehouse was done in a MOLAP environment, and the tools of analysis and interrogation of data were established with the aim to construct a decision-support tool that provides intuitive navigation through the available measures. The decision maker can rely on these indicators to monitor and control the air pollution phenomena in the county as well as when it comes to define short- and long-term actions. We have also developed the cartographic component of our application that makes possible the use of geographic and thematic tools of interrogation and analysis.

We are currently moving from the prototype presented in this paper to a more comprehensive air quality decision support system that will be help on monitoring air pollution over the country of Morocco. The latest will be enriched with realistic data and will rely on web service interfaces for offering various navigation options to end users.

8. REFERENCES

Agrawal R, et al, September 1995, Modeling Multidimensional Databases, *IBM Research Report, IBM Almaden Research Center*.

Bédard, Y.et al. 2006. Merging Hypermedia GIS with Spatial On-Line Analytical Processing: Towards Hypermedia SOLAP, in *Geographic Hypermedia: Concepts and Systems*, pp. 167-185.

Bimonte.S, et al, 2006. GeoCube, a Multidimensional Model and Navigation Operators Handling Complex Measures: Application in Spatial OLAP. *Actes de 4th Int. Conf. Advances in Information Systems, Izmir*, Turkey.

Chaudhuri S. and Dayal. U. 1997. An overview of data warehousing and OLAP technology. *SIGMOD Record*: 26 (1), pp 65-74.

Chimere, Chimere Specifications 2005, http://euler.lmd.polytechnique.fr/chimere/

Franklin, C. 1992, An Introduction to Geographic Information Systems*: Linking maps to databases*. Database, pp 13-21.

Kimball, R. 2002, The Data Warehouse Toolkit. *John Wiley and Sons, Inc., second edition*.

Mark Richards, et al. March 2006, Grid based analysis of air pollution data, *Ecological Modelling*, Volume 194 Issues 1–3 pp.274–286 Elsevier.

ProClarity, 2006. ProClarity Product Information, http://www.ProClarity.org/.

Proulx, M.et al, 2006, Exploration d'applications décisionnelles à la Direction des inventaires forestiers : *Développement d'un tableau de bord et d'une application Spatial OLAP (SOLAP), Rapport de recherche pour Direction des inventaires forestière- MRNPF, Mars*, 69p.

Rivest S. et al, 2001 Toward better support for spatial decision making: *Defining the characteristics of Spatial On-Line Analytical Processing, Geomatica*,Vol. 55, No. 4, pp. 539-555.

Taher Omran Ahmed, Maryvonne Miquel. 2005, Multidimensional Structures Dedicated to Continuous Spatiotemporal Phenomena. *BNCOD* : pp 29-40.

Vassiliadis P, 1998, Modeling Multidimensional Databases, Cubes and Cube Operations, *Proc. 10th Int. Conference on Statistical and Scientific Database Management (SSDBM)*, Capri, Italy, 1-6.